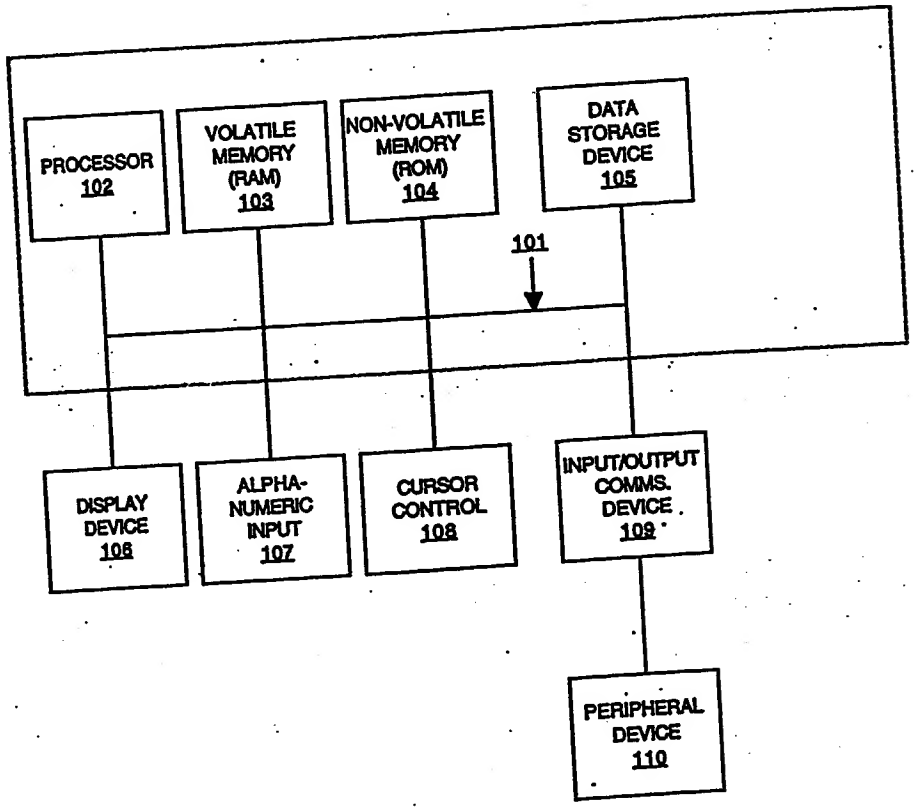
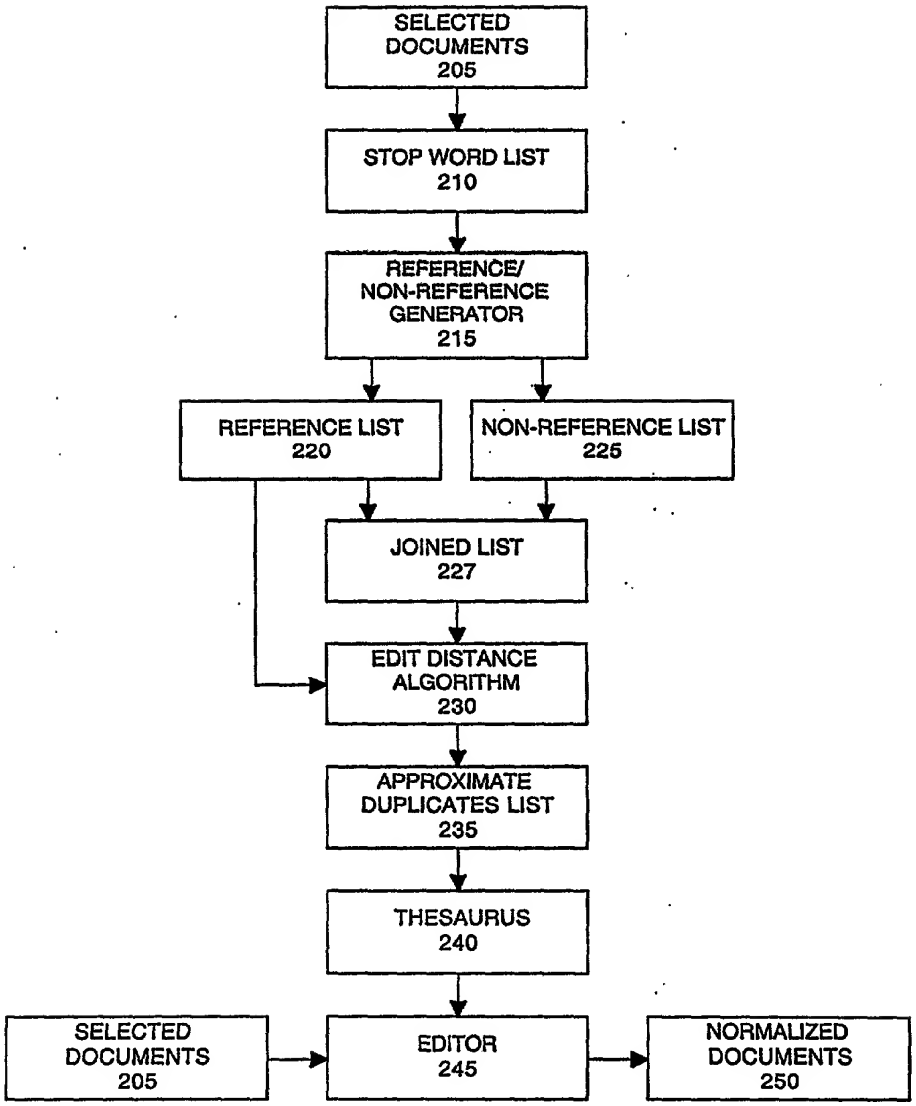
SCANNED, #

FIGURE 1



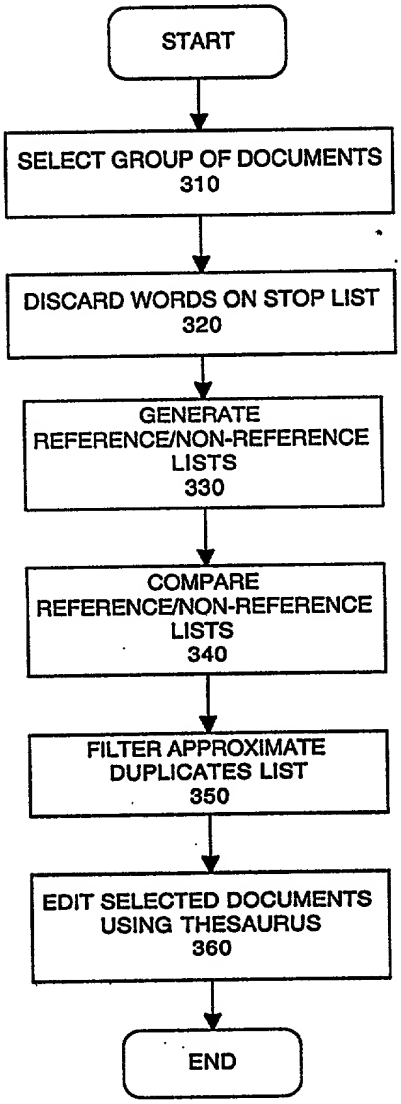
Patent Pending

FIGURE 2



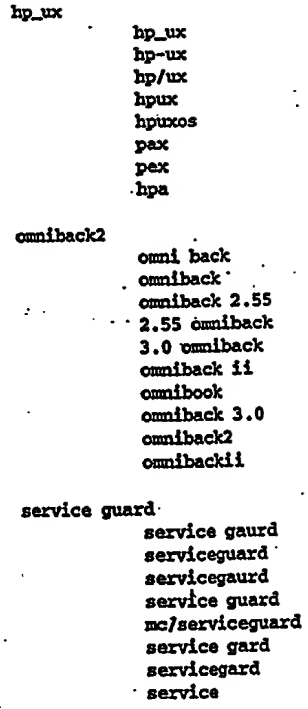
300

FIGURE 3


START
SELECT GROUP OF DOCUMENTS
310
DISCARD WORDS ON STOP LIST
320
GENERATE
REFERENCE/NON-REFERENCE
LISTS
330
COMPARE
REFERENCE/NON-REFERENCE
LISTS
340
FILTER APPROXIMATE
DUPLICATES LIST
350
EDIT SELECTED DOCUMENTS
USING THESAURUS
360
END

FOE120 09950660

FIGURE 4



DOCID: US 6950860 B1

FIGURE 5

Normalization Results

Number of documents normalized: 1568 out of 5000

WORDS	PHRASES	PER DOCUMENT		
		MIN	AVG	MAX
3385	88	1	2.2	71

TERM	REPLACEMENT TERM	DOC FREQ	TERM FREQ
10.2	10.20	50	63
os:10.20	10.20	58	69
version:10.20	10.20	121	171
ver:10.20	10.20	73	73
hpinstal	install	3	16
7install	install	0	0
patch(es	patch	21	40
xpatches	patch	4	5
megapatch	patch	11	13
patchset	patch	2	9
patch_id	patch	4	5
openned	opened	6	11
serv	server	0	0
hp_9000	9000	0	0